Title: Changes in Neural Processing After Multimodal Speech-Gesture Training in Patients With Schizophrenia

Authors: Lydia Riedl[1,2], Arne Nagels[3], Gebhard Sammer[4], Momoko Choudhury[1], Annika Nonnenmann[1], Anne Sütterlin[1], Chiara Feise[1], Maxi Haslach[1], Florian, Bitsch[1], Hildegard Kühne[5], Thomas S. Hartmann[1,2], Nina Shvetsova[5], Benjamin Straube[1,2]

Affiliations:

[1] Translational Neuroimaging Lab, Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Germany

[2] Center for Mind, Brain and Behavior (CMBB), Philipps-University Marburg and Justus Liebig University Giessen, Germany

[3] Department of English and Linguistics, Johannes-Gutenberg-University Mainz, Germany

[4] Department of Psychiatry and Psychotherapy, Justus-Liebig-University Giessen, Germany

[5] CVAI Group, Goethe-University Frankfurt, Germany

Abstract:

Background: Dysfunctional social communication is one of the most stable characteristics in patients with schizophrenia spectrum disorder that severely affects the patients' prognosis (e.g., regarding social life, work and life satisfaction). Interpreting abstract speech and integrating nonverbal information is particularly affected. According to former studies (Straube et al., 2013), aberrant neural activation in left inferior frontal gyrus (IFG) and superior/middle temporal regions (STS, MTG) could account to those dysfunctions.

Considering the difficulty to treat communication dysfunctions with usual intervention (Hegarty et al., 1994; Jääskeläinen et al., 2013), we investigated the possibility to improve co-verbal gesture processing in patients with schizophrenia by applying a multimodal speech-gesture (MSG) training.

Methods: In the MSG training, we offered eight sessions (60 min each) including perceptive and expressive tasks as well as meta-learning elements and transfer exercises to N=20 patients with schizophrenia spectrum disorder (SSD). Outcomes were changes in the neural processing of abstract speech-gesture information, which were measured before and after a no treatment period and again after an intensive MSG training period through functional Magnetic Resonance Imaging (fMRI).

Imaging data were collected with a 3 T whole body MRI system (SIEMENS MAGNETOM TrioTim syngo MR B17). During the fMRI measurements, videos of an actor expressing concrete (con) or abstract (abs) sentences accompanied by gesture (SG) were presented in a passive viewing task (see *Figure 1*).

Neural data were preprocessed with *fMRIPrep* 20.2.0rc0 (Esteban et al., 2018, 2020) (RRID:SCR_016216) including coarse gridding (2mm isotropic) and smoothing (kernel width FWHM=6.0mm).

Single subjects' voxel-wise BOLD activity was modeled by a General Linear Model (GLM) (Friston et al., 1995; Worsley & Friston, 1995). The single subject level parameter estimates were deployed for a full factorial analysis with *SPM12* [RRID: SCR_007037]. The main effect of condition were defined as contrasts of interest for the group level analysis, resulting in baseline contrasts for the two conditions of interest (absSG, conSG).

The contrast of interest focused on the interaction session X abstractness for the processing of multimodal video input. Whole-brain results were corrected for multiple comparisons on clusters covering at least 221 voxels, with a corrected threshold for rejection of the null hypothesis of $p < .01$ (thresholds were identified on the basis of a Monte Carlo simulation and applied to all functional analyses).

Results: Neural activation increased in the middle temporal gyrus for the processing of abstract multimodal content and decreased for the processing of concrete multimodal content, but only during the MSG training period and less during the no treatment period (see *Figure 2*). Improvement during training, self-report measures and ratings of relatives confirmed the MSG-related changes.

Conclusion: The comparison of the three measurements provides evidence for possible training-specific effects through comparing intra-individual repetition and training effects. Together, we provide first promising results of a novel multimodal speech-gesture training for the neural processing in patients with schizophrenia spectrum disorder.

Outlook: In order to investigate the underlying mechanisms of changes in neural processing of multimodal video input after an intensive MSG training, we currently perform analyses of audio-visual concept learning in the *Audio-Video Language Network* (AVLnet) (Rouditchenko et al., 2021), a self-supervised network that learns a shared audio-visual embedding space directly from raw video inputs. Through comparing AVLnet's learned representations with the MSG training induced changes in patients with schizophrenia spectrum disorder, we seek to investigate the implications for future therapy approaches.

Figures:

Figure 1:

Example of stimuli.

Illustration of the stimulus conditions:

(1) sentences with abstract content accompanied with gesture (absSG);

(2) sentences with concrete content accompanied with gesture (conSG).

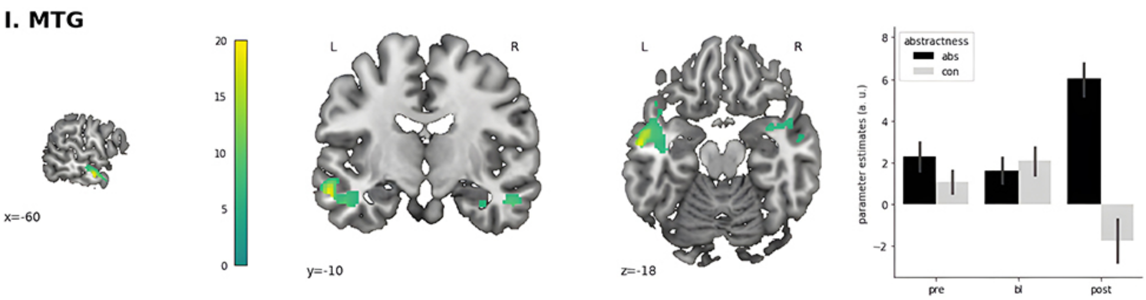

Figure 2: Interaction (F-test) of all three sessions X abstractness in patients with schizophrenia spectrum disorder. Activation clusters were thresholded at $p < 0.01$, with a minimum cluster size of 221 voxels, cluster level corrected at $p > 0.05$. Crosshair point at the left MTG [x = -60, y = -10, z = -18], right MTG [x = 52, y = -2, z = -30] and left CERCRU1 [x = -44, y = -58, z = -32].

CERCRU1 = Crus I of cerebellar hemisphere; MTG = middle temporal gyrus; L = left; R = right; abs = abstract multimodal condition (absSG); con = concrete multimodal condition (conSG); pre = session pre, first baseline measurement (before no treatment period); bl = session bl, second baseline measurement (after no treatment period and before MSG training); post = session post, last measurement (after MSG training).

References:

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLOS ONE*, *12*(9), e0184661. https://doi.org/10.1371/journal.pone.0184661

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2018). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Esteban, O., Markiewicz, C. J., Goncalves, M., DuPre, E., Kent, J. D., Salo, T., Ciric, R., Pinsard, B., Blair, R. W., Poldrack, R. A., & Gorgolewski, K. J. (2020). *fMRIPrep: A robust preprocessing pipeline for functional MRI*. Zenodo. https://doi.org/10.5281/zenodo.4252786

Friston, K. J., Holmes, A. P., Poline, J.-B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., & Turner, R. (1995). Analysis of fMRI Time-Series Revisited. *NeuroImage*, *2*(1), 45–53. https://doi.org/10.1006/nimg.1995.1007

Hegarty, J. D., Baldessarini, R. J., Tohen, M., Waternaux, C., & Oepen, G. (1994). One hundred years of schizophrenia: A meta-analysis of the outcome literature. *American Journal of Psychiatry*, *151*(10), 1409–1416.

Jääskeläinen, E., Juola, P., Hirvonen, N., McGrath, J. J., Saha, S., Isohanni, M., Veijola, J., & Miettunen, J. (2013). A Systematic Review and Meta-Analysis of Recovery in Schizophrenia. *Schizophrenia Bulletin*, *39*(6), 1296–1306. https://doi.org/10.1093/schbul/sbs130

Rouditchenko, A., Boggust, A., Harwath, D., Chen, B., Joshi, D., Thomas, S., Audhkhasi, K., Kuehne, H., Panda, R., Feris, R., Kingsbury, B., Picheny, M., Torralba, A., & Glass, J. (2021). AVLnet: Learning Audio-Visual Language Representations from Instructional Videos. *ArXiv:2006.09199 [Cs, Eess]*. http://arxiv.org/abs/2006.09199

Straube, B., Green, A., Sass, K., Kirner-Veselinovic, A., & Kircher, T. (2013). Neural integration of speech and gesture in schizophrenia: Evidence for differential processing of metaphoric gestures. *Human Brain Mapping*, *34*(7), 1696–1712. https://doi.org/10.1002/hbm.22015

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI Time-Series Revisited—Again. *NeuroImage*, *2*(3), 173–181. https://doi.org/10.1006/nimg.1995.1023

Categories:

Keywords: schizophrenia, psychiatry, fMRI, gesture, multimodal processing, neural processing, intervention, training, clinical trial, metaphor